Ecommerce Purchases

I use Fake Data about some purchases done through Amazon

all the data is fake and made up.

Import pandas and read in the Ecommerce Purchases csv file and set it to a DataFrame called ecom.

```
In[1]: import numpy as np
       import pandas as pd
```

```
In[2]: ecom = pd.read_csv("Ecommerce Purchases")
       ecom
```

```
Out[2]:
       Address  Lot  AM PM  Browser Info  Company  Credit Card  Exp Date  CC Security Code  CC Provider  Email  Job  IP Address  Language  Purchase Price
0  19029 Pace Camp Apt. 448  Kinshasha-by-the-Sea, MT 67...  PM  Opera/9.26.(Windows NT 5.1; sv-SE,en)  Martinez-Herman  6011509651125456  10/20  980  JCB 16 digit  johnhargrove@yahoo.com  Scientific productivity development  149.146.147.205  el  98.14
1  5974 Jasmine Spurs Suite 685  Lake Brittanyburgh, ... 2000  AM  Mozilla/5.0 (Windows NT 6.0; sl-SI; rv:1.9.0.7) Gecko/2009021910 Firefox/3.0.7  Gallegos, Smith and Pham  6759578661625  09/19  699  JCB 15 digit  amymiller@morales-harrison.com  Drilling engineer  15.180.41.51  fr  70.73
2  1790 Julia Fundamental Stravenue  Stacy, WA 95790  AM  Mozilla/5.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2.0.50727; .NET CLR 3.0.04506.30; .NET CLR 3.0.04506.648; .NET CLR 3.5.21022; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E; InfoPath.2; .NET4.0E; .NET CLR 3.5.30729; .NET CLR 3.0.30729; .NET CLR 2.0.50727; .NET CLR 1.1.4322; .NET CLR 3.5.21022)  Williams, Williams and Williams  6011795044437710  02/24  384  Discover  brent10@brown-robinson.info  Outdoor engineer  30.250.74.19  es  77.42
3  2953 Mexico Drive Suite 337  Kimberly, CA 90293-7574  PM  Opera/9.20.(Linux x86_64; en-US)  Brown, Wilson and Rodriguez  6011456652307988  10/25  679  Diners Club / Carte Blanche  christophermyers@yahoo.com  Fire artist  24.140.33.94  es  77.42
4  968 Cedarview Meadows Lookout  Jackson, CO 80477  AM  Mozilla/5.0 (Windows NT 5.2; Trident/4.0)  Parsons, Ramirez, Stone and Sharp  3425645032587851  03/22  838  JCB 16 digit  scott@wade-garner.com  Printmaker  29.72.157.114  fr  93.8
5  832 Colin Dam Suite 959  Emeraldhaven, TX 90847  AM  Mozilla/5.0 (Windows NT 6.1; Trident/5.0)  Kerr-Stewart  6011929690901240  10/26  207  JCB 16 digit  mwyatt@hotmail.com  Dye worker  121.133.168.51  fr  70.94
6  1442 Ashli Burg Apt. 594  Gonzalezburgh, WV 24801-5263  AM  Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6.8; rv:2.0.1) Gecko/20100101 Firefox/4.0.1  Anderson Group  6011397873026511  05/21  1  VISA 16 digit  tylerc3@gmail.com  Veterinary surgeon  156.210.0.254  el  77.42
9998  0789 Rodriguez Mill  Lake Ericbury, ND 90127-6...  PM  Mozilla/5.0 (Macintosh; Intel PowerPc Mac OS X 10_6_7; rv:1.9.2.20) Gecko/20110803 Firefox/3.6.20  Cook Inc  180003048629260  11/17  987  American Express  elizabethmoore@mail.net  Local government officer  55.78.16.143  es  70.73
9999  4874 Barrett Stravenue  Wilmington, DE 19810-5...  AM  Opera/9.27.(Windows NT 5.1; en-US,en)  Gonzalez, Cruz and Ortiz  4139729921773  02/19  302  JCB 15 digit  rachelfox@yahoo.com  Embroiderer clerical  179.119.188.199  el  77.42
```

10000 rows × 14 columns

Check the head of the DataFrame.

```
In[3]: ecom.head()
```

```
Out[3]:
   Address  Lot  AM PM  Browser Info  Company  Credit Card  Exp Date  CC Security Code  CC Provider  Email  Job  IP Address  Language  Purchase Price
0  19029 Pace Camp Apt. 448  Kinshasha-by-the-Sea, MT 67...  PM  Opera/9.26.(Windows NT 5.1; en)  Martinez-Herman  6011509651125456  10/20  980  JCB 16 digit  johnhargrove@yahoo.com  Scientific productivity development  149.146.147.205  el  98.14
1  5974 Jasmine Spurs Suite 685  Lake Brittanyburgh, ... 2000  AM  Mozilla/5.0 (Windows NT 6.0; sl-SI; rv:1.9.0.7) Gecko/2009021910 Firefox/3.0.7  Gallegos, Smith and Pham  6759578661625  09/19  699  JCB 15 digit  amymiller@morales-harrison.com  Drilling engineer  15.180.41.51  fr  70.73
2  1790 Julia Fundamental Stravenue  Stacy, WA 95790  AM  Mozilla/5.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2.0.50727; .NET CLR 3.0.04506.30; .NET CLR 3.0.04506.648; .NET CLR 3.5.21022; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E; InfoPath.2; .NET4.0E; .NET CLR 3.5.30729; .NET CLR 3.0.30729; .NET CLR 2.0.50727; .NET CLR 1.1.4322; .NET CLR 3.5.21022)  Williams, Williams and Williams  6011795044437710  02/24  384  Discover  brent10@brown-robinson.info  Outdoor engineer  30.250.74.19  es  77.42
3  2953 Mexico Drive Suite 337  Kimberly, CA 90293-7574  PM  Opera/9.20.(Linux x86_64; en-US)  Brown, Wilson and Rodriguez  6011456652307988  10/25  679  Diners Club / Carte Blanche  christophermyers@yahoo.com  Fire artist  24.140.33.94  es  77.42
4  968 Cedarview Meadows Lookout  Jackson, CO 80477  AM  Mozilla/5.0 (Windows NT 5.2; Trident/4.0)  Parsons, Ramirez, Stone and Sharp  3425645032587851  03/22  838  JCB 16 digit  scott@wade-garner.com  Printmaker  29.72.157.114  fr  93.8
```

How many rows and columns are there?

```
In[4]: ecom.info()
       ecom.shape
```

```
Out[4]:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
#   Column            Non-Null Count  Dtype
---  ------            --------------  -----
0   Address           10000 non-null  object
1   Lot               10000 non-null  object
2   AM PM             10000 non-null  object
3   Browser Info      10000 non-null  object
4   Company           10000 non-null  object
5   Credit Card       10000 non-null  object
6   CC Exp Date       10000 non-null  object
7   CC Security Code  10000 non-null  int64
8   CC Provider       10000 non-null  object
9   Email             10000 non-null  object
10  Job               10000 non-null  object
11  IP Address        10000 non-null  object
12  Language          10000 non-null  object
13  Purchase Price    10000 non-null  float64
dtypes: float64(1), int64(1), object(12)
memory usage: 1.1+ MB

(10000, 14)
```

What is the average Purchase Price?

```
In[5]: ecom['Purchase Price'].mean()
```

```
Out[5]:
50.347302999999975
```

The average Purchase Price is **$50.34**

What were the highest and lowest purchase prices?

```
In[6]: ecom['Purchase Price'].max()
```

```
Out[6]:
99.99
```

```
In[7]: ecom['Purchase Price'].min()
```

```
Out[7]:
0.0
```

The highest is **99.99 dollar** The lowest is **0 dollar**

How many people have English 'en' as their Language of choice on the website?

```
In[8]: ecom['Language'].unique()
       ecom[ecom['Language'] == 'en'].shape
```

```
Out[8]:
array(['el', 'fr', 'de', 'es', 'ru', 'pl', 'sh', 'no', 'it'], dtype=object)

(10000,14)
(1000,14)
```

```
In[9]: ecom.shape
       ecom[ecom['Language'] == 'en'].value_counts()
```

```
Out[9]:
(10000, 14)
Address                                            Lot  AM or PM  Browser Info
IP Address  Language  Purchase Price  Company  Credit Card  CC Exp Date  CC Security Code  CC Provider  Email  Job
0000 Jared Shoals Apt. 244\nFredericksburg, IL 12450  55 Z2  PM  Mozilla/5.0 (compatible; MSIE 6.0; Windows CE; Trident/3.1)
Sisco-Ferguson  6011797667891785  03/17  837  Discover  bellcrawford@rogers.org  Biomedical engineer
91.108.150.74  en  50.19  Mccarthy, Ortiz and Ramirez  3572139000276109  09/21  109  MasterCard  andrewd42@yu-miranda.org  Tourism officer
7694 Monica Rapids\nSuite Moore, MT 24925  K3 H7  PM  Mozilla/5.0 (compatible; MSIE 6.0; Windows NT 5.2; Trident/3.1)
Ortiz-Ramirez  4017594883281883  10/23  48  VISA 16 digit  williamsmelissa@yahoo.net  Paramedic
26.71.11.132  en  50.19  Ortiz-Ramirez  3572139000276109  09/21  109  MasterCard  andrewd42@yu-miranda.org  Tourism officer
75351 Linda Mountain\nHinckleyburgh, IL 32111-6353  37 G4  PM  Mozilla/5.0 (iPod; U; CPU iPhone OS 4\_2\_1 like Mac OS X; fr-FR) AppleWebKit/532.9 (KHTML, Gecko) Version/4.0.5 Mobile/8B117 Safari/6531.22.7  Gecko/2010-07-10 03:04:59 01
Anderson Group  6011397873026511  06/22  613  JCB 16 digit  tylerc3@gmail.com  Veterinary surgeon
156.210.0.254  el  77.42  Stewart-Ortiz  3088764471665460  04/17  873  JCB 15 digit  ericmartinez@hotmail.com  Textile engineer
7852 Sandra Glens Suite 854\nGreenport, MS 53396  88 A7  PM  Mozilla/5.0 (iPod; U; CPU iPhone OS 4\_2\_1 like Mac OS X; fr-FR) AppleWebKit/532.9 (KHTML, Gecko) Version/4.0.5 Mobile/8B117 Safari/6531.22.7  Gecko/2013-08-08 18:20:11 01
Anderson Group  6011397873026511  06/21  679  Diners Club / Carte Blanche  mwyatt@hotmail.com  Dye worker
121.133.168.51  fr  70.94  Stewart-Ortiz  3088764471665460  04/17  873  JCB 15 digit  ericmartinez@hotmail.com  Textile engineer
276 Woodard Mill Apt. 505\nLawsonport, ND 68186  97 Fw  AM  Mozilla/5.0 (Windows NT 6.1; it-IT; rv:1.9.1.9) Gecko/2010-08-07 10:56:5
4 Firefox/3.6.4  Lawson Ltd  4762929610649152  01/19  295  VISA 16 digit  ericmartinez@hotmail.com  Control and instrumentation engineer  172.103.53.71  el  77.42
277 Kelly Forges\nSuite 836\nPhoenixville, CA 70404  7H  AM  Mozilla/5.0 (compatible; MSIE 8.0; Windows NT 6.0; Trident/4.1)
Yang-Kelly  6011575762593513  09/21  109  MasterCard  andrewd42@yu-miranda.org  Tourism officer
Tourism officer  172.103.53.71  el  77.42  Ortiz-Ramirez  3572139000276109  09/21  109  MasterCard  andrewd42@yu-miranda.org  Tourism officer
27870 Estelle Pike\nEast William, MC 77922  9G  PM  Mozilla/5.0 (compatible; MSIE 8.0; Windows NT 6.2; Trident/5.1)
Quinn and Sons  6011892194777413  10/22  362  Discover  dawsonchristian@gmail.com  Town planner
Town planner  91.108.150.74  es  50.19  Dixon-Dixon  6011892194777413  10/22  362  Discover  dawsonchristian@gmail.com  Town planner
30122 Stewart Shoals Suite 937\nLeachburgh, NY 97567  D5 AD  PM  Opera/9.26.(Linux i686; en-US) Presto/2.9.171 Version/10.00
Rosales Group  3088843073864285  03/23  84  VISA 16 digit  lijamesc@hotmail.com  Toxicologist
Toxicologist  172.103.53.71  es  50.19  Ortiz-Ramirez  3572139000276109  09/21  109  MasterCard  andrewd42@yu-miranda.org  Tourism officer
unit 8867 box 0871\APO AE 46354  98 NT  AM  Opera/9.80.(X11; Linux x86_64; en-US) Presto/2.9.168 Version/11.00
Cooper, James and Perkins  6041235781  04/26  455  Discover  steevefreeman@yahoo.com  Sports administrator
172.103.53.71  es  50.19  Ortiz-Ramirez  3572139000276109  09/21  109  MasterCard  andrewd42@yu-miranda.org  Tourism officer
Length: 1000, dtype: int64
```

```
In[10]: ecom[ecom['Language'] == 'en'].count()
```

```
Out[10]:
Address             1000
Lot                 1000
AM PM               1000
Browser Info        1000
Company             1000
Credit Card         1000
CC Exp Date         1000
CC Security Code    1000
CC Provider         1000
Email               1000
Job                 1000
IP Address          1000
Language            1000
Purchase Price      1000
dtype: int64
```

The number of users who have selected English 'en' as their preferred language on the website is **1000**

How many people have the job title of "Lawyer" ?

```
In[11]: ecom[ecom['Job'] == "Lawyer"].info()
```

```
Out[11]:
<class 'pandas.core.frame.DataFrame'>
Int64Index: 30 entries, 1911 to 9877
Data columns (total 14 columns):
#   Column            Non-Null Count  Dtype
---  ------            --------------  -----
0   Address           30 non-null     object
1   Lot               30 non-null     object
2   AM PM             30 non-null     object
3   Browser Info      30 non-null     object
4   Company           30 non-null     object
5   Credit Card       30 non-null     object
6   CC Exp Date       30 non-null     object
7   CC Security Code  30 non-null     int64
8   CC Provider       30 non-null     object
9   Email             30 non-null     object
10  Job               30 non-null     object
11  IP Address        30 non-null     object
12  Language          30 non-null     object
13  Purchase Price    30 non-null     float64
dtypes: float64(1), int64(1), object(12)
memory usage: 3.5+ KB
```

The number of individuals with the job title 'Lawyer' is **30**

How many people made the purchase during the AM and how many people made the purchase during PM ?

(Hint: Check out value_counts())

```
In[12]: ecom[ecom['AM or PM'] == 'AM'].shape
        ecom[ecom['AM or PM'] == 'PM'].shape
```

```
Out[12]:
(5068, 14)
(4932, 14)
```

The number of people who made purchases during the AM is **4932**, and the number of people who made purchases during the PM is **5068**

What are the 5 most common Job Titles?

```
In[13]: ecom['Job'].value_counts().sort_values(ascending=False).head(5)
```

```
Out[13]:
Interior and spatial designer    31
Lawyer                           30
Purchasing manager               29
Purchasing manager               27
Designer, jewellery              27
Name: Job, dtype: int64
```

The 5 most common Job Titles

1. Interior and spatial designer 31
2. Lawyer 30
3. Social researcher 28
4. Purchasing manager 27
5. Designer, jewellery 27

Someone made a purchase that came from Lot: "90 WT" , what was the Purchase Price for this transaction?

```
In[14]: ecom[ecom['Lot'] == "90 WT"]["Purchase Price"]
```

```
Out[14]:
513    75.1
Name: Purchase Price, dtype: float64
```

The purchase price for the transaction that came from Lot 90 WT is **$75.1**

What is the email of the person with the following Credit Card Number: 4926535242672853

```
In[15]: ecom[ecom["Credit Card"] == 4926535242672853]["Email"]
```

```
Out[15]:
1234    bond41@williams-garza.com
Name: Email, dtype: object
```

The email associated with the credit card number 4926535242672853 is **bond41@williams-garza.com**

How many people have American Express as their Credit Card Provider and made a purchase above $95?

```
In[16]: ecom[("C:\Users\jose\AppData\Local\Temp\ipykernel_11896\#2922600.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
ecom[ecom['CC Provider'] == 'American Express']]ecom['Purchase Price'] > 95).count()
```

```
Out[16]:
Address             39
Lot                 39
AM or PM            39
Browser Info        39
Company             39
Credit Card         39
CC Exp Date         39
CC Security Code    39
CC Provider         39
Email               39
Job                 39
IP Address          39
Language            39
Purchase Price      39
dtype: int64
```

The number of individuals who have American Express as their Credit Card Provider and made a purchase above $95 is **39**

How many people have a credit card that expires in 2025?

```
In[17]: ecom[ecom[
        ecom[ecom[
```

```
Out[17]:
   Address  Lot  AM PM  Browser Info  Company  Credit Card  Exp Date  CC Security Code  CC Provider  Email  Job  IP Address  Language  Purchase Price
0  19029 Pace Camp Apt. 448  Kinshasha-by-the-Sea, MT 67...  PM  Opera/9.26.(Windows NT 5.1; en)  Martinez-Herman  6011509651125456  10/20  980  JCB 16 digit  johnhargrove@yahoo.com  Scientific productivity development  149.146.147.205  el  98.14
1  5974 Jasmine Spurs Suite 685  Lake Brittanyburgh, ... 2000  AM  Mozilla/5.0 (Windows NT 6.0; sl-SI; rv:1.9.0.7) Gecko/2009021910 Firefox/3.0.7  Gallegos, Smith and Pham  6759578661625  09/19  699  JCB 15 digit  amymiller@morales-harrison.com  Drilling engineer  15.180.41.51  fr  70.73
2  1790 Julia Fundamental Stravenue  Stacy, WA 95790  AM  Mozilla/5.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322; .NET CLR 2.0.50727; .NET CLR 3.0.04506.30; .NET CLR 3.0.04506.648; .NET CLR 3.5.21022; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; .NET4.0C; .NET4.0E; InfoPath.2; .NET4.0E; .NET CLR 3.5.30729; .NET CLR 3.0.30729; .NET CLR 2.0.50727; .NET CLR 1.1.4322; .NET CLR 3.5.21022)  Williams, Williams and Williams  6011795044437710  02/24  384  Discover  brent10@brown-robinson.info  Outdoor engineer  30.250.74.19  es  77.42
3  2953 Mexico Drive Suite 337  Kimberly, CA 90293-7574  PM  Opera/9.20.(Linux x86_64; en-US)  Brown, Wilson and Rodriguez  6011456652307988  10/25  679  Diners Club / Carte Blanche  christophermyers@yahoo.com  Fire artist  24.140.33.94  es  77.42
4  968 Cedarview Meadows Lookout  Jackson, CO 80477  AM  Mozilla/5.0 (Windows NT 5.2; Trident/4.0)  Parsons, Ramirez, Stone and Sharp  3425645032587851  03/22  838  JCB 16 digit  scott@wade-garner.com  Printmaker  29.72.157.114  fr  93.8
```

```
In[18]: ecom[ecom['CC Exp Date'] == '10/25']
```

```
Out[18]:
   Address  Lot  AM PM  Browser Info  Company  Credit Card  Exp Date  CC Security Code  CC Provider  Email  Job  IP Address  Language  Purchase Price
3  2953 Mexico Drive Suite 337  Kimberly, CA 90293-7574  PM  Opera/9.20.(Linux x86_64; en-US)  Brown, Wilson and Rodriguez  6011456652307988  10/25  679  Diners Club / Carte Blanche  christophermyers@yahoo.com  Fire artist  24.140.33.94  es  77.42
```

```
In[19]: ecom[ecom['CC Exp Date'] == 10/25]
```

```
Out[19]:
1033
```

The number of individuals with a credit card that expires in 2025 is **1033**

What are the top 5 most popular email providers/hosts (e.g. gmail.com, yahoo.com, etc...)

```
In[20]: ecom["Email"]
```

```
Out[20]:
0              johnhargrove@yahoo.com
1              amymiller@morales-harrison.com
2              brent10@brown-robinson.info
3              christophermyers@yahoo.com
4              scott@wade-garner.com
                      ...
9995           lewisbrittany@hotmail.com
9996           ericmartinez@hotmail.com
9997           mwyatt@hotmail.com
9998           elizabethmoore@mail.net
9999           rachelfox@yahoo.com
Name: Email, Length: 10000, dtype: object
```

```
In[21]: ecom[ecom['Email'].apply(lambda x: x.split('@')[1])].value_counts().head()
```

```
Out[21]:
hotmail.com  1638
yahoo.com    1616
gmail.com    1305
smith.com    417
williams.com 207
Name: Email, dtype: int64
```

The top 5 most popular email providers/hosts, based on user data, are

1. hotmail.com 1638 2. yahoo.com 1616
2. gmail.com 1305
3. smith.com 417
4. williams.com 17

Great Job!